Successive Approximation Methods in Undiscounted Stochastic Games

AWI FEDERGRUEN
Columbia University, New York, New York
(Received June 1977; accepted September 1979)

This paper considers undiscounted two-person, zero-sum sequential games with finite state and action spaces. Under conditions that guarantee the existence of stationary optimal strategies, we present two successive approximation methods for finding the optimal gain rate, a solution to the optimality equation, and for any $\epsilon > 0$, ϵ-optimal policies for both players.

THIS PAPER considers two-person, zero-sum stochastic games with finite state space $\Omega = \{1, \cdots, N\}$ and in each state $i \in \Omega$, two finite sets $K(i)$ and $L(i)$ of actions available to player 1 and 2, respectively. The state of the system is observed at equidistant epochs. When the system is observed to be in state i, the two players choose an action, or a randomization of actions out of $K(i)$ and $L(i)$, respectively. When the actions $k \in K(i)$, $l \in L(i)$ are chosen in state i, then $P_{ij}^{k,l} \geq 0$ denotes the probability that state j is the next state to be observed ($\sum_{j=1}^{N} P_{ij}^{k,l} = 1$) and $q_i^{k,l}$ is the one-step expected reward earned by player 1 from player 2.

If the payoffs are discounted at the interest rate $r > 0$, the stochastic game is called the *r-discounted game.* The existence of a value and of stationary optimal policies in the r-discounted game goes essentially back to Shapley [22]; in addition it is easily verified that value-iteration converges to the value of the game, in view of the value-iteration operator being a contraction mapping on E^N, the N-dimensional Euclidean space.

In the *undiscounted* version of the game, i.e., when the long run average return per unit time is the criterion to be considered, one or both players may fail to have optimal stationary policies, as follows from an example in Gillette [11]. Both for this model and for the case of more general state and action spaces, recurrency conditions with respect to the transition probability matrices (*tpm's*) associated with the stationary policies have been obtained under which the existence of a stationary pair of equilibrium policies (*AEP*) is guaranteed (see Federgruen [7], Hoffman and Karp [13], Rogers [18], Sobel [23], and Stern [24]).

So far, very little attention has been paid to the actual computation of both the asymptotic average value g^* and of a solution v^* to the average

Operations Research
Vol. 28, No. 3, Part II, May–June 1980
0030-364X/80/2803-0794 $01.25

return optimality equation (see Section 1), under conditions that guarantee the existence of a stationary AEP.

In view of the fact that the value of (both the discounted and undiscounted version of) the game does not necessarily lie within the same ordered field as the parameters of the problem (see Bewley and Kohlberg [2]) we cannot expect to find a finite algorithm in the sense that it involves a finite number of field-operations.

Two algorithms were given by Hoffman and Karp [13] and Pollatschek and Avi-Itzhak [17]. It was shown that the first algorithm converges to a stationary AEP, if the tpm of each pure stationary policy pair is unichained and has no transient states. Although the second algorithm seems to compare favorably with the first one, as far as net running time and the required number of iterations is concerned, it is still unknown under which conditions its convergence is guaranteed.

In this paper, we provide two successive approximation methods for locating optimal policies for both players. In both algorithms, we obtain in addition at each step of the iteration procedure, upper and lower bounds for the asymptotic average value which converge to the latter as the number of iteration steps tends to infinity.

The first algorithm is an adaptation of a "modified" value-iteration method as introduced by Bather [1] and as generalized by Hordijk and Tijms [14]. Its convergence is guaranteed whenever condition (H1) below is satisfied.

(H1): (a) a stationary AEP exists.
(b) the asymptotic average value g^* is independent of the initial state of the system.

The second algorithm is based upon the more elementary value-iteration method, and may successfully be applied whenever condition (H2) below holds:

(H2): the tpm of each of the pure stationary policy pairs is unichained.

Note that (H2) $\Rightarrow$ (H1) (see e.g. [7], Theorem 3). Under (H2) we obtain in addition lower and upper bounds for the fixed point v^* of the optimality equation which in this case is unique up to a multiple of $\mathbf{1}$, where $\mathbf{1}$ is the N-vector with all components unity.

At each step of the procedure, both methods merely require the solution of N relatively small linear programs (the size of which is determined by the number of actions in $K(i)$ and $L(i)$, $i \in \Omega$). Especially for large scale systems, i.e., when $N \gg 1$, this compares favorably with the techniques used in [13] and [17] which require at each step of the procedure the solution of a system of at least N equations.

One might wish to extend these methods to the more general stochastic renewal games-model (*SRG*) in which the state of the system is not

necessarily observed at equidistant epochs. In the SRG-model we make the more general assumption that when the actions $k \in K(i)$ and $l \in L(i)$ are chosen in state i, the period of time until the next observation of state is a random variable with finite and positive expectation $T_i^{k,l}$ $(i \in \Omega;\ k \in K(i),\ l \in L(i))$. In the one-player case, this model reduces to a Markov renewal program (*MRP*) (see Denardo and Fox [6] and Jewell [15]). The value-iteration methods that were developed for the discrete-time or Markov decision problem (*MDP*) case, could easily be extended to the MRP-model, due to a data-transformation which was introduced in Schweitzer [20] and which turns every undiscounted MRP into an equivalent MDP. In the two-player case, the analog of this data-transformation, will generally fail to work. The only exception is provided by the case where the expected holding times $T_i^{k,l}$ $(i \in \Omega,\ k \in K(i),\ l \in L(i))$ satisfy the *separability* assumption:

$$\text{(SEP)}\ T_i^{k,l} = \alpha_i^k \beta_i^l \quad \text{with} \quad \alpha_i^k,\ \beta_i^l > 0; \qquad i \in \Omega,\ k \in K(i),\ l \in L(i). \tag{1}$$

This will be shown in the appendix, Section 4. (1) holds e.g. in case the transition time between two successive observations of the state of the system merely depends upon the current state, possibly in combination with the action chosen by one of the two players. Establishing an efficient algorithm for the *general* SRG-case remains, however, an outstanding problem.

In Sections 2 and 3 we present our two methods, and in Section 1 we give some notation and preliminaries.

1. NOTATION AND PRELIMINARIES

For any finite set A, let $\| A \|$ denote the number of elements it contains. If $A = [A_{ij}]$ is a matrix, let $|A| = \max_{i,j} |A_{ij}|$ and let val A indicate the value of the corresponding matrix game. Note that for any pair of matrices A, B of equal dimension:

$$|\operatorname{val} A - \operatorname{val} B| \leq |A - B|. \tag{2}$$

(Let (x^A, y^A) and (x^B, y^B) be equilibrium pairs of actions in the matrix games A and B; the $\min_{i,j}(A_{ij} - B_{ij}) \leq x^B(A - B)y^A = x^B A y^A - x^B B y^A \leq \operatorname{val} A - \operatorname{val} B \leq x^A A y^B - x^A B y^B = x^A(A - B)y^B \leq \max_{i,j}(A_{ij} - B_{ij})$.) For all $i \in \Omega$, and any set of numbers $\{c_i^{k,l} \mid k \in K(i),\ l \in L(i)\}$, $[c_i^{k,l}]$ denotes the $\| K(i) \| \times \| L(i) \|$ matrix, the (k, l)-th entry of which is $c_i^{k,l}$.

For all $r > 0$, let $V(r)$ denote the vector, the ith component of which denotes the value of the r-discounted game, with initial state $i \in \Omega$. Bewley and Kohlberg [2] recently showed that $V(r)$ may be expressed as a real fractional power or Puiseux series in r, for all interest rates r, sufficiently close to 0. More specifically, there exists an integer $M \geq 1$ such that:

$$V(r) = g^*/r + \sum_{k=-\infty}^{M-1} a^{(k)} r^{-k/M} \tag{3}$$

We call g^* the *asymptotic average value vector.*

A player's policy is a rule which prescribes for each stage $t = 1, 2, \cdots$ which (randomized) action to choose, depending upon the current state and the entire history of the game up to that stage. A policy is said to be stationary if it prescribes actions which depend merely upon the current state of the system, regardless of the stage of the game, and its history up to this stage. Note that a stationary strategy $f(h)$ for player 1(2) is characterized by a tableau $[f_{ik}]([h_{il}])$ satisfying $f_{ik} \geq 0$ and $\sum_{k \in K(i)} f_{ik} = 1 (h_{ik} \geq 0$ and $\sum_{l \in L(i)} h_{il} = 1)$, where $f_{ik}(h_{il})$ is the probability that the kth (lth) alternative in $K(i)(L(i))$ is chosen when entering state $i \in \Omega$. We let $\Phi(\Psi)$ denote the set of all stationary policies for player 1(2). We associate with each pair $(f, h) \in \Phi \times \Psi$ *a* N-component reward vector $q(f, h)$, a holding rate vector $T(f, h)$ and a stochastic matrix $P(f, h)$:

$$\begin{aligned} q(f, h)_i &= \sum_{k \in K(i)} \sum_{l \in L(i)} f_{ik} q_i^{k,l} h_{il}; \qquad i \in \Omega \\ T(f, h)_i &= \sum_{k \in K(i)} \sum_{l \in L(i)} f_{ik} T_i^{k,l} h_{il}; \qquad i \in \Omega \\ P(f, h)_{ij} &= \sum_{k \in K(i)} \sum_{l \in L(i)} f_{ik} P_{ij}^{k,l} h_{il}; \qquad i, j \in \Omega \end{aligned} \tag{4}$$

Finally we define for any pair $(f, h) \in \Phi \times \Psi$ the stochastic matrix $\Pi(f, h)$ as the Cesaro limit of the sequence $\{P^n(f, h)\}_{n=1}^{\infty}$. Since we employ the long run average return per unit time criterion, we evaluate any pair (ϕ, ψ) of (possibly nonstationary) policies for players 1 and 2, by considering the gain rate vector $g(\phi, \psi)$:

$$g(\phi, \psi)_i = \lim_{n \to \infty} (E_{\phi,\psi} \sum_{l=1}^{n} \rho_l) / (E_{\phi,\psi} \sum_{l=1}^{n} \tau_l); \qquad i \in \Omega \tag{5}$$

where $\rho_n(\tau_n)$ denotes the payoff to player 1 (the length of the period) in between the $(n-1)$-st and the nth observation of state. $E_{\phi,\psi}$ indicates the expectation given the players' policies ϕ and ψ. A number of equivalent criteria have been formulated in [4]. $g(\phi, \psi)_i$ equals the limit as $t \to \infty$, of the expected average cost incurred by time t (see Theorem 7.5 in Ross [19]).

A pair of stationary policies (ϕ^*, ψ^*) is called an AEP, if and only if for every policy pair (ϕ, ψ)

$$g(\phi, \psi^*)_i \leq g(\phi^*, \psi^*)_i \leq g(\phi^*, \psi)_i, \qquad \text{for all} \quad i \in \Omega. \tag{6}$$

One easily verifies (see e.g. [2] and [8] that if (f^*, h^*) is a stationary AEP, $g(f^*, h^*) = g^*$.

In [8], we showed that a pair of optimality equations arises when considering the average return per unit time criterion, and we investigated the interdependences between the existence of a stationary AEP and a solution to this pair of optimality equations.

In the case where $g_i^* = \langle g^* \rangle$, $i \in \Omega$, i.e., when the asymptotic average value is independent of the initial state (see condition (H1)), this pair of optimality equations reduces to the single equation:

$$v_i + g = \text{val}[q_i^{k,l} + \textstyle\sum_j P_{ij}^{k,l} v_j], \qquad i \in \Omega. \tag{7}$$

LEMMA 1 (see Corollary 2.5 in [8]). *Assume that* $g_i^* = \langle g^* \rangle$, $i \in \Omega$. *Then the following statements are equivalent*: (*I*) $a^{(k)} = 0$, $k = 1, \cdots, M-1$ (*see* (3)), (*II*) *there exists a stationary AEP, and* (*III*) (7) *has a solution pair* (g, v).

In addition, under either one of (I), (II) or (III), any solution pair (g, v) has $g = g^*$, and any policy pair $(f^*, h^*) \in \Phi \times \Psi$ which satisfies the equation (7), i.e., which attains the N equilibria in (7) simultaneously for some solution pair (g, v), is an AEP.

Finally, let $V = \{v \in E^N | (g^*, v) \text{ satisfies } (7)\}$.

2. A MODIFIED VALUE-ITERATION TECHNIQUE

Throughout this section, we assume (H1) to hold, which implies in view of Lemma 1, the existence of a solution pair (g^*, v^*) to the optimality equation (7). We investigate the behavior of the scheme

$$y(n+1)_i = \text{val}[q_i^{k,l} - g^* + (1 + r_n)^{-1} \textstyle\sum_j P_{ij}^{k,l} y(n)_j], \qquad i \in \Omega \tag{8}$$

with $y(0)$ a given N-vector. It follows from the proof of Theorem 2.3 in [8] that $\lim_{n\to\infty} y(n) = a^{(0)}$, provided that the sequence $\{r_n\}_{n=1}^{\infty}$ satisfies

$$(1 - r_n) \cdots (1 - r_2) \to 0, \qquad \text{as} \quad n \to \infty \tag{9}$$

$$\textstyle\sum_{j=2}^{n} (1 - r_n) \cdots (1 - r_j) | r_j^{1/M} - r_{j-1}^{1/M} | \to 0, \qquad \text{as} \quad n \to \infty \tag{10}$$

where $(a^{(0)}, g^*)$ is a solution pair to the optimality equation (7).

LEMMA 2. *Conditions* (9) *and* (10) *are satisfied for any choice*: $r_n = n^{-b}$ *with* $0 < b \leq 1$.

Proof. Note using the mean value theorem that $n^b - (n-1)^b \leq 1$ for all $n = 1, 2, \cdots$ and use this inequality in order to verify that:

$$(1 - r_n) \cdots (1 - r_2) = \frac{(n^b - 1)}{n^b} \frac{((n-1)^b - 1)}{(n-1)^b} \cdots \frac{(2^b - 1)}{2^b} \leq \frac{2^b - 1}{n^b}$$

which proves (9). Next we apply the mean value theorem to verify that

$$\textstyle\sum_{j=2}^{n} (1 - r_n) \cdots (1 - r_j) | r_j^{M^{-1}} - r_{j-1}^{M^{-1}} |$$
$$\leq bM^{-1} \sum_{j=2}^{n} \frac{(n^b - 1)}{n^b} \cdots \frac{(j^b - 1)}{j^b} (j-1)^{-bM^{-1}-1}$$

$$\leq bM^{-1}n^{-b}\sum_{j=2}^{n} j^{b(1-M^{-1})-1}$$

$$\leq bM^{-1}n^{-b}\int_{1}^{n} x^{b(1-M^{-1})-1}\,dx$$

$$= \begin{cases} bn^{-b}\ln(n), & \text{if } M = 1 \\ (M-1)^{-1}n^{-bM^{-1}}, & \text{otherwise} \end{cases}$$

which proves (10).

Remark 1. For the MDP—i.e., one player—case, Lemma 2 indicates a larger range of permitted values for b, than the one that was obtained in [14] (p. 206 remark) using a different analysis.

Observe that the sequence $\{y(n)\}_{n=1}^{\infty}$ cannot be computed in view of g^* being unknown. We circumvent this numerical difficulty as in White [25], i.e., we define the sequences $\{\hat{y}(n)\}_{n=1}^{\infty}$ and $\{G(n)\}_{n=1}^{\infty}$ by:

$$\begin{aligned} \hat{y}(n+1)_i &= y(n+1)_i - y(n+1)_N \\ &= \text{val}[q_i^{k,l} + (1+r_n)^{-1}\textstyle\sum_j P_{ij}^{k,l}\hat{y}(n)_j] - G(n+1); \qquad (11) \\ &\qquad\qquad i \in \Omega \qquad n = 0, 1, 2, \cdots \end{aligned}$$

$$\begin{aligned} G(n+1) &= \text{val}[q_N^{k,l} + (1+r_n)^{-1}\textstyle\sum_j P_{Nj}^{k,l}\hat{y}(n)_j]; \\ &\qquad\qquad i \in \Omega; \qquad n = 0, 1, 2, \cdots \end{aligned} \qquad (12)$$

where $\hat{y}(0)_i = y(0)_i - y(0)_N$; $i \in \Omega$.

THEOREM 1. *For all* $n = 1, 2, \cdots$ *let*:

$$\begin{aligned} D(n+1) &= \min_i\{\text{val}[q_i^{k,l} + (1+r_n)^{-1}\textstyle\sum_j P_{ij}^{k,l}\hat{y}(n)_j] \\ &\qquad - (1+r_n)^{-1}\hat{y}(n)_i\} \\ U(n+1) &= \max_i\{\text{val}[q_i^{k,l} + (1+r_n)^{-1}\textstyle\sum_j P_{ij}^{k,l}\hat{y}(n)_j] \\ &\qquad - (1+r_n)^{-1}\hat{y}(n)_i\}. \end{aligned} \qquad (13)$$

(a) *Let* (f^*, h^*) *be a stationary AEP and for any* $n = 1, 2, \cdots$ *let* $(f_n, h_n) \in \Phi \times \Psi$ *be any pair of policies which attain the N equilibria to the right of* (11) *simultaneously. Then*

(1) $D(n) \leq G(n) \leq U(n), \qquad n = 1, 2, \cdots$

(2) $D(n+1) \leq g(f_n, h^*)_i \leq g^* \leq g(f^*, h_n)_i \leq U(n+1); \qquad i \in \Omega$

(b) *If* $\{r_n\}_{n=1}^{\infty}$ *satisfies* (9) *and* (10), *then*:

$$\lim_{n\to\infty} D(n) = \lim_{n\to\infty} G(n) = \lim_{n\to\infty} U(n) = g^*$$

$$\lim_{n\to\infty} \hat{y}(n) = a^{(0)} - \langle a_N^{(0)}\rangle \underline{1} \in V$$

where $\mathbf{1}$ *is the vector all of whose components are unity.*

Proof. (a) (1) Note from (11) that $\hat{y}(n)_N = 0$ for all $n = 0, 1, 2, \cdots$, hence $D(n) \leq \text{val}[q_N^{k,l} + (1 + r_n)^{-1} \sum_j P_{Nj}^{k,l} \hat{y}(n)_j] = G(n) \leq U(n)$.

(2) The inner inequalities are immediate from the fact that (f^*, h^*) is a stationary AEP. We next prove the most left inequality $D(n + 1) \leq g(f_n, h^*)$, the proof of $g(f^*, h_n) \leq U(n + 1)$ being analogous. Note that for all $i \in \Omega$:

$$D(n + 1) + (1 + r_n)^{-1} \hat{y}(n)_i \leq \text{val}[q_i^{k,l} + (1 + r_n)^{-1} \textstyle\sum_j P_{ij}^{k,l} \hat{y}(n)_j]$$
$$\leq q(f_n, h^*)_i + (1 + r_n)^{-1} P(f_n, h^*) \hat{y}(n)_i,$$

and multiply both sides of this inequality by $\Pi(f_n, h^*)_i \geq 0$ and sum on $i \in \Omega$.

(b) Recall that $\lim_{n\to\infty} y(n) = a^{(0)} \in V$. Next we observe that if $v \in V$ then so is $v + c\mathbf{1}$ for all scalars c. Hence,

$$\lim_{n\to\infty} \hat{y}(n) = \lim_{n\to\infty} y(n) - \langle y(n)_N \rangle \mathbf{1} = a^{(0)} - \langle a_N^{(0)} \rangle \mathbf{1} \in V$$

This in combination with the fact that the "val"-operator is (Lipschitz) continuous (see (2)) imply using (7):

$$\lim_{n\to\infty} D(n) = \min_i \{\text{val}[q_i^{k,l} + \textstyle\sum_j P_{ij}^{k,l} a_j^{(0)}] - a_i^{(0)}\}$$
$$= \min_i g^* = \langle g^* \rangle = \max_i \{\text{val}[q_i^{k,l} + \textstyle\sum_j P_{ij}^{k,l} a_j^{(0)}] - a_i^{(0)}\}$$
$$= \lim_{n\to\infty} U(n)$$

which together with part (a)(1) completes the proof of (b).

Remark 2. When taking $r_n = n^{-b}$ for some b, with $0 < b \leq 1$, the approach to the limits in part (b) of the above theorem, exhibits a convergence rate which is of the order

$$\begin{cases} 0(n^{-b} \ln n), & \text{if } M = 1 \\ 0(n^{-bM^{-1}}), & \text{otherwise} \end{cases}$$

as follows from the proof of Lemma 2 and Theorem 2.3 in [8]. We note that the bounds for g^* in part (a)(2) generalize the bounds Odoni [16] and Hastings [12] obtained for the MDP-case.

We summarize this section by specifying an algorithm which approximates g^*, as well as a solution $v \in V$, and which finds for any $\epsilon > 0$, ϵ-optimal policies for both players:

Algorithm 1

Step 0. Fix a sequence $\{r_n\}_{n=1}^{\infty}$ satisfying (9) and (10), e.g., take $r_n = n^{-b}$ with $0 < b \leq 1$. Set $n = 0$, fix $y(0) \in E^N$ and $\epsilon > 0$.

Step 1. Calculate $\hat{y}(n + 1)$, $D(n + 1)$, $G(n + 1)$ and $U(n + 1)$ from $\hat{y}(n)$, using (11), (12) and (13).

Step 2. If $U(n+1) - D(n+1) < \epsilon$ determine a stationary policy pair (f_n, h_n) which attains the N equilibria to the right of (11) simultaneously; use $f_n(h_n)$ as an ϵ-optimal policy for player 1(2); $G(n+1)$ as an ϵ-approximation for g^* and $\hat{y}(n+1)$ as an approximation for a solution $v \in V$. Otherwise increment n by one and return to Step 1.

3. VALUE-ITERATION; A SUFFICIENT CONDITION FOR CONVERGENCE

In this section, we discuss the asymptotic behavior of the sequence:

$$v(n+1)_i = Qv(n)_i, \qquad i \in \Omega \tag{14}$$

where the operator Q: $E^N \to E^N$ is defined by $Qx_i = \text{val}[q_i^{k,l} + \sum_j P_{ij}^{k,l} x_j]$, $i \in \Omega$, and where $v(0) \in E^N$ is a given N-vector.

Note that the Q operator is monotonic, and satisfies the basic properties:

$$Q(x + c\mathbf{1}) = Qx + c\mathbf{1} \quad \text{for all scalars} \quad c; \qquad x \in E^N \tag{15}$$

$$(x-y)_{\min} \le (Qx - Qy)_{\min} \le (Qx - Qy)_{\max} \le (x-y)_{\max}; \qquad x, y \in E^N \tag{16}$$

where (16) is easily verified by applying the Q-operator to both sides of the inequalities $y + (x-y)_{\min}\mathbf{1} \le x$ and $x \le y + (x-y)_{\max}\mathbf{1}$, using its monotonicity as well as (15).

Note that $v(n)_i$ may be interpreted as the value of the n-stage game when starting in state i and given some final amount $v(0)_j$ is earned by player 1 from player 2, when ending up in state j.

Whereas we still have $\lim_{n\to\infty} v(n)/n = g^*$ (see Bewley and Kohlberg [2], Theorem 3.2) the difference $\{v(n) - ng^*\}_{n=1}^{\infty}$ does not need to be bounded, in sharp contrast to the behavior in the one player-model (see Brown [5], Theorem 4.3).

In fact, Bewley and Kohlberg [3] proved the existence of a number $B > 0$ and a Puiseux series in n,

$$W(n) = ng^* + \sum_{k=-\infty}^{\hat{M}-1} b^{(k)} n^{k/\hat{M}}$$

such that $|v(n) - W(n)| < B \log(n+1)$, $n = 1, 2, \cdots$.

LEMMA 3. *$\{v(n) - ng^*\}_{n=1}^{\infty}$ is bounded under condition (H1).*

Proof. Note from Lemma 1, that (H1) implies the existence of a solution $v \in V$. Next, use (2) in order to conclude that:

$$|v(n) - ng^* - v| \le |\text{val}[q_i^{k,l} + \textstyle\sum_j P_{ij}^{k,l} v(n-1)_j] - \text{val}[q_i^{k,l} + \textstyle\sum_j P_{ij}^{k,l}(v + (n-1)g^*)]_j| \le |v(n-1) - (n-1)g^* - v|.$$

It is known from Markov decision theory that even in case $\{v(n) - ng^*\}_{n=1}^{\infty}$ is bounded, the sequence may fail to converge if some of the tpm's of the pure stationary policy pairs happen to be periodic (in [21], Schweitzer and Federgruen obtained for the MDP-case the necessary and sufficient condition for $\{v(n) - ng^*\}_{n=1}^{\infty}$ to converge for all $v(0) \in E^N$).

In this section we first apply a data-transformation which turns our stochastic game into an equivalent one, in the sense that the two stochastic games have the same gain rate vector for any stationary pair of policies, and hence the same asymptotic average value vector and the same set of stationary AEPs. We next analyze the behavior of (14) under condition (H2) which is a stronger version of (H1) (see the introduction). The data-transformation is analogous to the one employed in Schweitzer [20]:

$$\tilde{q}_i^{k,l} = q_i^{k,l}; \qquad i \in \Omega,\ k \in K(i),\ l \in L(i), \tag{17}$$

$$\tilde{P}_{ij}^{k,l} = \tau(P_{ij}^{k,l} - \delta_{ij}) + \delta_{ij}; \qquad i, j \in \Omega,\ k \in K(i),\ l \in L(i) \tag{18}$$

where $0 < \tau < 1$ and where δ_{ij} represents the Kronecker-delta function. Verify that $\tilde{P}_{ij}^{k,l} \geq 0$, $\sum_j \tilde{P}_{ij}^{k,l} = 1$, with the gain rate vector of each pair of policies in $\Phi \times \Psi$ remaining unaltered by the data-transformation.

In addition, each of the tpm's of the stationary policy pairs in the transformed model, has a positive *diagonal*, which obviously implies aperiodicity. Let $\tilde{Q}$ be the value iteration operator in the transformed model, and define $\tilde{V}$ as the solution set of its optimality equation (7). Finally let $\{\tilde{v}(n)\}_{n=1}^{\infty} = \{\tilde{Q}^n v(0)\}_{n=1}^{\infty}$.

LEMMA 4. (a) $\tilde{V} = \{v \in E^N \mid \tau v \in V\}$, *and* (b) *If* $(f^*, h^*) \in \Phi \times \Psi$ *satisfy the optimality equation* (7) *in the original* [*transformed*] *model for some* $v \in V$ [$\tilde{v} \in \tilde{V}$], *then they will satisfy the optimality equation in the transformed* [*original*] *model for* $\tau^{-1}v[\tau\tilde{v}]$ *as well.*

Proof. Consider an arbitrary two-person zero-sum game $[c_i^{k,l}]$ for some $i \in \Omega$. Then for any constant a and positive number b:

$$\begin{aligned} &\text{(i)}\ \operatorname{val}[c_i^{k,l} + a] = \operatorname{val}[c_i^{k,l}] + a \\ &\text{(ii)}\ \operatorname{val}[bc_i^{k,l}] = b \operatorname{val}[c_i^{k,l}] \end{aligned} \tag{19}$$

with the set of equilibrium pairs of action remaining unaltered, both by the translation, and by the (positive) scalar multiplication. Use (19) while rewriting (7) as

$$0 = \operatorname{val}[q_i^{k,l} - g^* + \textstyle\sum_j (P_{ij}^{k,l} - \delta_{ij})v_j], \qquad i \in \Omega \quad \text{or}$$

$$0 = \operatorname{val}[q_i^{k,l} - g^* + \textstyle\sum_j \tau(P_{ij}^{k,l} - \delta_{ij})(\tau^{-1}v_j)], \qquad i \in \Omega$$

Next, we restrict the analysis of the $\tilde{Q}$-operator on the N-1-dimensional

subspace $\bar{E}^N = \{x \in E^N \mid x_N = 0\}$, considering the following reduction $\bar{Q}$ of the $\tilde{Q}$-operator:

$$\bar{Q}\colon \bar{E}^N \to \bar{E}^N\colon x \to \tilde{Q}x - < \tilde{Q}x_N > \mathbf{1}.$$

Accordingly, define $\bar{v}(n)_i = \tilde{v}(n)_i - \tilde{v}(n)_N = \bar{Q}\bar{v}(n-1)_i$, $i \in \Omega$. (Note the similarity with the reduction in White [25] and of $\{y(n)\}_{n=1}^{\infty}$ to $\{\hat{y}(n)\}_{n=1}^{\infty}$ in (11)).

We call a function $\Lambda(x)$ on a vector space X, a *Lyapunov* function with *origin* $x^* \in X$, if:

$$\begin{aligned} &(1)\ \Lambda(x) \quad \text{is continuous on} \quad X \\ &(2)\ \Lambda(x) \geq 0 \quad \text{and} \quad \Lambda(x) = 0 \Leftrightarrow x = x^*. \end{aligned} \tag{20}$$

We have not been able to obtain a straightforward analysis of the behavior of $\{v(n)\}_{n=1}^{\infty}$ or $\{\bar{v}(n)\}_{n=1}^{\infty}$. However, the study of difference equations of the type (14) may be greatly facilitated with the help of Lyapunov functions, as is shown by the following lemma.

LEMMA 5. *Let $\Lambda(x)$ be a Lyapunov function on a vector space X, with origin x^*. For $n = 2, 3, \cdots$ let A^n denote the n-fold application of an operator $A\colon X \to X$ i.e., $A^{n+1}x = A(A^n x)$. Then,*

$$\begin{aligned} &\lim_{n\to\infty} A^n x = x^*, \quad \textit{for all } x \in X, \textit{ if} \\ &(1)\ \Lambda(Ax) \leq \Lambda(x), \quad \textit{for all } x \in X \\ &(2)\ \textit{there exists an integer } J \geq 1 \textit{ such that } \Lambda(A^J x) < \Lambda(x), \\ &\quad \textit{for all } x \neq x^*. \end{aligned} \tag{21}$$

Lemma 5 is an immediate adaptation of Theorem 10.4 in Zangwill [26]. In the context of Markov decision theory, the use of Lyapunov functions, and in particular of Lemma 5, was first pointed out in [9].

Now, under (H2), the solution to the optimality equation (7) is unique up to a multiple of $\mathbf{1}$, as was shown in [6], Theorem 3.1, i.e., on $\bar{E}^N$ there exists a *unique* solution $v^* \in \tilde{V}$.

We next observe that both $\Lambda_1(x)$ and $\Lambda_2(x)$ are Lypapunov functions on $\bar{E}^N$ with v^* as origin, where

$$\begin{aligned} \Lambda_1(x) &= \| x - v^* \|_d \\ \Lambda_2(x) &= \| \bar{Q}x - x \|_d = \| \tilde{Q}x - x \|_d, \end{aligned}$$

with $\| x \|_d = \max_i x_i - \min_i x_i$ (see Bather [1]). $\Lambda_1(x)$ obviously satisfies both conditions in (20); $\Lambda_2(x) \geq 0$ is immediate as well, its continuity on $\bar{E}^N$ follows from the continuity of the "val"-operator (see (1)) and $\| Qx - x \|_d = 0 \Leftrightarrow$ there exists a scalar g, such that $Qx - x = \langle g \rangle \mathbf{1} \Leftrightarrow x \in \tilde{V} \cap \bar{E}^N \Leftrightarrow x = v^*$).

Note that $\Lambda_2(x)$ has the advantage of being computable at each point $x \in \bar{E}^N$.

We next recall that in the transformed model, and as a consequence of assumption (H2) the tpm's of all stationary policy pairs are unichained, and in addition have all diagonal entries strictly positive. In Theorem 4 of [10], Federgruen, Schweitzer and Tijms showed that this implies the following "scrambling-type" condition for all pairs of N-tuples of pure policy pairs $\{(f_1, h_1); \cdots; (f_N, h_N)\}$ and $\{(f_1', h_1'); \cdots; (f_N', h_N')\}$:

$$\sum_{j=1}^{N} \min[P(f_N, h_N) \cdots P(f_1, h_1)_{i_1 j}; P(f_N', h_N') \cdots P(f_1', h_1')_{i_2 j}] > 0$$

$$\text{for all} \quad i_1 \neq i_2. \quad (22)$$

Observe that for all $i_1, i_2 \in \Omega$ the expression to the left of the above inequality is a continuous function on $[X_{l=1}^{N} \Phi \times \Psi]^2$ which can be embedded as a compact subset of a Euclidean space. Hence there exists a uniform scrambling coefficient $\alpha > 0$, such that

$$\sum_{j=1}^{N} \min[P(f_N, h_N) \cdots P(f_1, h_1)_{i_1 j}; P(f_N', h_N') \cdots P(f_1', h_1')_{i_2 j}] > \alpha$$

$$\text{for all} \quad i_1 \neq i_2; \quad (f_l, h_l) \quad \text{and} \quad (f_l', h_l') \in \Phi \times \Psi \ (l = 1, \cdots, N). \quad (23)$$

This enables us to prove the convergence of $\{\bar{v}(n)\}_{n=1}^{\infty}$ under (H2). Let $d(n+1) = [\tilde{Q}\bar{v}(n) - \bar{v}(n)]_{\min}$ and $u(n+1) = [\tilde{Q}\bar{v}(n) - \bar{v}(n)]_{\max}$ for all $n = 0, 1, \cdots$. Define $g(n+1) = [\tilde{Q}\bar{v}(n)]_N$.

THEOREM 2. *Assume a data transformation* (17)-(18) *is employed.* (a) *Both* $\Lambda_1(x)$ *and* $\Lambda_2(x)$ *satisfy* (21) *with* $J = N$; *hence* $\lim_{n\to\infty} \bar{v}(n) = v^*$ *for all* $v(0) \in E^N$; (b) $d(n) \leq d(n+1) \leq g(n+1) \leq u(n+1) \leq u(n)$ *for all* $n = 1, 2, \cdots$ $\lim_{n\to\infty} d(n) = \lim_{n\to\infty} g(n) = \lim_{n\to\infty} u(n) = g^*$; *and* (c) *let* $(f^*, h^*) \in \Phi \times \Psi$ *be an AEP, and for all* $n = 1, 2, \cdots$ *let* $(f_n, h_n) \in \Phi \times \Psi$ *be any pair of policies which attain the N equilibria to the right of* (14) *simultaneously. Then* $l(n+1) \leq g(f_n, h^*) \leq g^* \leq g(f^*, h_n) \leq u(n+1)$.

Proof. (a) We merely show that $\Lambda_1(x)$ satisfies (21), the proof for $\Lambda_2(x)$ being analogous. Use (20) to verify that $\Lambda_1(\tilde{Q}x) = \| \tilde{Q}x - v^* \|_d = \| \tilde{Q}x - \tilde{Q}v^* \|_d \leq \| x - v^* \|_d = \Lambda_1(x)$. Next we obtain part (2) of condition (21) by showing that:

$$\Lambda_1(\tilde{Q}^N x) = \Lambda_1(\tilde{Q}^N x) < (1 - \alpha)\Lambda_1(x), \qquad \text{for all } x \in X \quad (24)$$

where the proof of (24) goes along lines with the proof of Theorem 5 in [8], using (23).

(b) The proof of $d(n+1) \leq m(n+1) \leq u(n+1)$ is analogous to the proof of part (a)(1) in Theorem 1; next note that $d(n+1) = [\tilde{Q}\bar{v}(n) - \bar{v}(n)]_{\min} = [\tilde{Q}(\tilde{Q}\bar{v}(n-1)) - \tilde{Q}\bar{v}(n-1)]_{\min} \leq [\tilde{Q}\bar{v}(n-1) - \bar{v}(n-1)]_{\min} =$

$d(n)$, where the inequality part follows from (16). The monotonicity of $\{u(n)\}_{n=1}^{\infty}$ is shown in complete analogy.

(c) See proof of Theorem 1 part (a)(2).

Observe that (24) is stronger than condition (20)(2), since the latter does not require the existence of some integer $J \geq 1$, for which $\sup_{x \in X} \Lambda(A^J x)/\Lambda(x) < 1$.

In fact (24) shows that the approach to all of the limits in parts (a) and (b) of the above theorem exhibits a *geometric* rate of convergence, which is considerably better than the rates we obtained in Section 2, for Algorithm 1 (see Remark 2). In this particular case, it is even possible to show (along lines with the proof of Theorem 5 in [10]) that $\bar{Q}$ is an N-step contraction mapping on $\bar{E}^N$, i.e., $\| \bar{Q}^N x - \bar{Q}^N y \|_d \leq (1 - \alpha) \| x - y \|_d$, for all $x, y \in E^N$ and the latter leads to the following bounds on v^*:

$$\bar{v}(nN + r)_i - \alpha^{-1}(1 - \alpha)^N \| v(N) - v(0) \|_d \leq v_i^* \leq$$
$$\bar{v}(nN + r)_i + \alpha^{-1}(1 - \alpha)^N \| v(N) - v(0) \|_d; \tag{25}$$
$$i \in \Omega; \qquad n = 1, 2, \cdots; \qquad r = 0, \cdots N - 1$$

(for a proof see [10], Theorem 6 part (a)).

Finally we conclude this section by specifying as in Section 2 an algorithm which approximates g^* as well as some $v \in V$, and for any $\epsilon > 0$, ϵ-optimal policies for both players:

Algorithm 2

Step 0. Fix $0 < \tau < 1$ and transform the stochastic game into an equivalent one with $(\tilde{q}_i^{k,l}; \tilde{P}_{ij}^{k,l})$ as the parameter set using the transformation formulas (17) and (18). Set $n = 0$; fix $v(0) \in E^N$ and $\epsilon > 0$.

Step 1 and Step 2. As in Algorithm 1, merely replacing $\hat{y}(n)$, $D(n)$, $G(n)$, $U(n)$ by $\bar{v}(n)$, $d(n)$, $g(n)$, and $u(n)$; $n = 1, 2, \cdots$.

Note that in this case (25) may be used as a stopping criterion for getting ϵ-approximations for v^*.

4. APPENDIX: ON REDUCING UNDISCOUNTED SRGs TO EQUIVALENT UNDISCOUNTED STOCHASTIC GAMES

In the introduction, we pointed out that in the *one-player* case, successive approximation methods for Markov renewal programs could be obtained by transforming the MRP-model into an equivalent undiscounted MDP-model. When trying to obtain a similar reduction for the SRG-case, thereby establishing an algorithm to solve the undiscounted version of this game, it is tempting to consider the natural extension of the data-transformation that is used in the *one*-player case (see [20]):

$$\tilde{q}_i^{k,l} = q_i^{k,l}/T_i^{k,l}; \qquad i \in \Omega; \qquad k \in K(i), \qquad l \in L(i)$$
$$\tilde{P}_{ij}^{k,l} = \delta_{ij} + (\tau/T_i^{k,l})[P_{ij}^{k,l} - \delta_{ij}]; \qquad i,j \in \Omega; \qquad k \in K(i); \qquad l \in L(i) \tag{26}$$
$$\tilde{T}_i^{k,l} = 1; \qquad i \in \Omega; \qquad k \in K(i); \qquad l \in L(i).$$

with

$$0 < \tau \leq \min\{1/(1 - P_{ii}^{k,l}) \mid i \in \Omega, \quad k \in K(i), \quad l \in L(i) \quad \text{with} \quad P_{ii}^{k,l} < 1\}. \tag{27}$$

Note that as a consequence of (27) $\tilde{P}_{ij}^{k,l} \geq 0$ and $\sum_j \tilde{P}_{ij}^{k,l} = 1$ for all $i, j \in \Omega$, $k \in K(i)$, $l \in L(i)$ such that it is possible to define a (related) discrete-time stochastic game which has $\{\tilde{q}_i^{k,l}\}$ as its one-step expected rewards and $\{\tilde{P}_{ij}^{k,l}\}$ as its set of one-step transition probabilities.

We recall from (2.14) in [8] that, under (H1), the solution of our SRG-model reduces to the problem of finding a vector $v \in E^N$ that satisfies the functional equation (see also (7) and Corollary 2.5 in [8]):

$$v_i = \text{val}[q_i^{k,l} - g^* T_i^{k,l} + \textstyle\sum_j P_{ij}^{k,l} v_j]; \qquad i \in \Omega. \tag{28}$$

Lemma 6 below shows that the above proposed reduction method works, in case the holding times $T_i^{k,l}$ satisfy the separability assumption (SEP) in (1). Let V denote the set of solutions to (28) and let $\tilde{V}$ be the set of solutions to the optimality equation in the transformed model. Likewise, all other quantities of interest in the transformed model will be marked off by a $(\sim)$-symbol.

LEMMA 6. *Suppose (H1) and (SEP) in (1) hold. For each pair of policies $f \in \Phi$ and $h \in \Psi$ define $\tilde{f}, \hat{f} \in \Phi$ and $\tilde{h}, \hat{h} \in \Psi$ by:*

$$\tilde{f}_{ik} = f_{ik}\alpha_i^k/\textstyle\sum_{r \in K(i)} f_{ir}\alpha_i^r; \qquad i \in \Omega, \qquad k \in K(i)$$
$$\hat{f}_{ik} = f_{ik}(\alpha_i^k)^{-1}/\textstyle\sum_{r \in K(i)} f_{ir}(\alpha_i^r)^{-1}; \qquad i \in \Omega, \quad k \in K(i) \tag{29}$$
$$\tilde{h}_{il} = h_{il}\beta_i^l/\textstyle\sum_{r \in L(i)} h_{ir}\beta_i^r; \qquad i \in \Omega, \qquad l \in L(i)$$
$$\hat{h}_{il} = h_{il}(\beta_i^l)^{-1}/\textstyle\sum_{r \in L(i)} h_{ir}(\beta_i^r)^{-1}; \qquad i \in \Omega, \qquad l \in L(i).$$

Then,

(a) $$g(f, h) = \tilde{g}(\tilde{f}, \tilde{h}) \quad \text{and} \quad \tilde{g}(f, h) = g(\hat{f}, \hat{h})$$

i.e., if (f, h) is an AEP in the original [transformed] model, then $(\tilde{f}, \tilde{h})$ $[(\hat{f}, \hat{h})]$ is an AEP in the transformed [original] model. In other words, there exists a computationally tractible one-to-one correspondence between the sets of stationary AEPs in the two models.

(b) $$\tilde{V} = \{v \in E^N \mid \tau v \in V\}.$$

Proof. We first consider an arbitrary matrix game $[c_i^{k,l}]$ for some $i \in \Omega$, in relationship with its "transformed" version $[c_i^{k,l}/T_i^{k,l}]$. Assuming that $\text{val}[c_i^{k,l}] = 0$, we prove the following two properties

$$\text{val}[c_i^{k,l}/T_i^{k,l}] = 0. \tag{30}$$

If $x^* \in \mathcal{K}(i)$ is an optimal action in the original {transformed} matrix game, then

$$\begin{aligned} \tilde{x} = [x_k{}^* \alpha_i^k / \textstyle\sum_r x_r{}^* \alpha_i^r]_{k \in K(i)} \\ \cdot \{\text{and } \hat{x} = [x_k{}^* (\alpha_i^k)^{-1} / \textstyle\sum_r x_r{}^* (\alpha_i^r)^{-1}]_{k \in K(i)}\} \end{aligned} \tag{31}$$

is an optimal action in the transformed {original} model. A similar one-to-one correspondence exists between the sets of optimal actions for player 2.

First, let $K(i) = \{x \in E^{\|K(i)\|} \mid x \geq 0, \sum_{h=1}^{\|K(i)\|} x_h = 1\}$ denote the set of all randomized actions available to player 1 in state i. Similarly $L(i) = \{y \in E^{\|L(i)\|} \mid y \geq 0, \sum_{l=1}^{\|L(i)\|} y_l = 1\}$ indicates the set of all randomized actions available to player 2 in state $i \in \Omega$.

Part (b) then follows by rewriting (28) in a homogeneous way, i.e., $0 = \text{val}[q_i^{k,l} - g^* T_i^{k,l} + \sum_j \tau(P_{ij}^{k,l} - \delta_{ij})(\tau^{-1} v)_j]$, $i \in \Omega$, invoking (30). The proof of part (a) follows from (31) and the observation that in the system

$$\begin{aligned} 0 &= [P(f, h) - I]g \\ 0 &= \{q(f, h)_i - g_i T(f, h)_i + [P(f, h) - I]v_i\}, \qquad i \in \Omega \end{aligned} \tag{32}$$

the g-part is uniquely determined as $g(f, h)$.

This leaves us with the proof of (30) and (31). Let (x^*, y^*) be a pair of equilibrium actions in the original matrix game. Then, for all $y \in \mathcal{L}(i)$:

$$\begin{aligned} &\textstyle\sum_k \sum_l \tilde{x}_k \, (c_i^{k,l} / \alpha_i^k \beta_i^l) y_l \\ &\quad = ((\textstyle\sum_r y_r \beta_i^r) / (\textstyle\sum_r x_r{}^* \alpha_i^r)) \textstyle\sum_k \sum_l x_k{}^* c_i^{k,l} [y_l \beta_i^l / \textstyle\sum_r y_r \beta_i^r] \geq 0 \end{aligned} \tag{33}$$

where the inequality follows from x^* being optimal in the original game. Likewise, with $\tilde{y} = [y_l{}^* \beta_i^l / \sum_r y_r{}^* \beta_i^r]_{l \in L(i)}$ we obtain

$$\textstyle\sum_k \sum_l x_k (c_i^{k,l} / \alpha_i^k \beta_i^l) \tilde{y}_l \leq 0 \quad \text{for all} \quad x \in \mathcal{K}(i) \tag{34}$$

such that (30) and (31) follow from the combination of (33) and (34).

We conclude that g^*, $v \in V$ and ϵ-optimal policies for both players can be computed by applying Algorithm 1 *under* (*H*1), *or Algorithm* 2 *under* (*H*2) *to the transformed model and by exploiting the one-to-one correspondences exhibited by Lemma* 6. *Note in addition, that by choosing τ strictly less than the upperbound in* (27) *the transformation in step* 0 *of Algorithm* 2 *becomes superfluous.*

The above described reduction fails, if the expected holding times fail to satisfy (SEP). This is due to (30) and (31) breaking down in general,

examples of which can easily be constructed. As a consequence, establishing an algorithm for the general SRG-case remains an outstanding problem.

REFERENCES

1. J. Bather, "Optimal Decision Procedures for Finite Markov Chains, Part II," *Adv. Appl. Prob.* **5,** 521–540 (1973).
2. T. Bewley and E. Kohlberg, "The Asymptotic Theory of Stochastic Games," *Math. Opns. Res.* **1,** 197–208 (1976).
3. T. Bewley and E. Kohlberg, "The Asymptotic Solution of a Recursive Equation Arising in Stochastic Games," *Math. Opns. Res.* **1,** 321–336 (1976).
4. T. Bewley and E. Kohlberg, "On Stochastic Games with Stationary Optimal Strategies," *Math. Opns. Res.* **3,** 104–126 (1978).
5. B. Brown, "On the Iterative Method of Dynamic Programming on a Finite State Space Discrete Time Markov Process," *Ann. Math. Stat.* **36,** 1279–1285 (1965).
6. E. Denardo and B. Fox, "Matichain Markov Renewal Programs," *SIAM J. Appl. Math.* **26,** 468–487 (1968).
7. A. Federgruen, "On N-person Stochastic Games with Denumerable State Space," *Adv. Appl Prob.* **10,** 452–472 (1978).
8. A. Federgruen, "On the Functional Equations in Undiscounted and Sensitive Discounted Stochastic Games," Math. Center Report BW 73/77, 1977.
9. A. Federgruen and P. J. Schweitzer, "On the Use of Lyapunov Functions in Markov Decision Theory" (forthcoming).
10. A. Federgruen, P. J. Schweitzer and H. C. Tijms, "Contraction Mappings Underlying Undiscounted Markov Decision Problems," *J. Math. Anal. Appl.* **65,** 711–730 (1978).
11. D. Gillette, "Stochastic Games with Zero Stop Probabilities," in *Contributions to the Theory of Games, Vol. III*, pp. 179–188, M. Dresher et al. (eds.), Princeton University Press, Princeton, N. J., 1957.
12. N. Hastings, "Bounds on the Gain of a Markov Decision Process," *Opns. Res.* **19,** 240–244 (1971).
13. A. Hoffman and R. Karp, "On Non-terminating Stochastic Games," *Mgmt. Sci.* **12,** 359–370 (1966).
14. A. Hordijk and H. Tijms, "A Modified Form of the Iterative Method of Dynamic Programming," *Ann. Stat.* **3,** 203–208 (1975).
15. W. Jewell, "Markov Renewal Programming," *Opns. Res.* **11,** 938–971 (1963).
16. A. Odoni, "On Finding the Maximal Gain for Markov Decision Processes," *Opns. Res.* **17,** 857–860 (1969).
17. M. Pollatschek and B. Avi-Itzhak, "Algorithms for Stochastic Games with Geometrical Interpretation," *Mgmt. Sci.* **15,** 399–415 (1969).
18. P. Rogers, "Nonzero-sum Stochastic Games," Report ORC 69-8, Operations Research Center, University of California, Berkeley, 1969.
19. S. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970.

20. P. J. SCHWEITZER, "Iterative Solution of the Functional Equations of Undiscounted Markov Renewal Programming," *J. Math. Anal. Appl.* **34,** 495–501 (1971).
21. P. J. SCHWEITZER AND A. FEDERGRUEN, "The Asymptotic Behavior of Undiscounted Value Iteration in Markov Decision Problems," *Math. Opns. Res.* **2,** 360–381 (1978).
22. L. SHAPLEY, "Stochastic Games," *Proc. Natl. Acad. Sci. U.S.A.* **39,** 1095–1100 (1953).
23. M. SOBEL, "Noncooperative Stochastic Games," *Ann. Math. Stat.* **42,** 1930–1935 (1971).
24. M. STERN, "On Stochastic Games with Limiting Average Payoff," Ph.D. dissertation, Department of Mathematics, University of Illinois, Chicago, 1975.
25. D. WHITE, "Dynamic Programming, Markov Chains and the Method of Successive Approximations," *J. Math. Anal. Appl.* **6,** 373–376 (1963).
26. W. ZANGWILL, *Nonlinear Programming, a Unified Approach*, Prentice-Hall, Englewood Cliffs, N. J., 1969.